Epigraphy Edit-a-thon
Editing chronological and geographic data in ancient inscriptions

April 20-22, 2016
Alexander von Humboldt Lehrstuhl für Digital Humanities
Universität Leipzig

Round Table Report

Chiara Palladino
(Università di Bari / Universität Leipzig)

Summary

General questions:

> *We need to think about how to integrate new systems with "old" museum and archival systems which are nevertheless still used for conservation* - people who take care of inscriptions need to feel welcome in order to be involved in a way that they can understand. We should start from the big collections so that minor collections feel that they can participate in the "prestigious group".

Problems of copyright and access to images: *Libraries are the new academic publishers.*
Cf. "Die dauerhafte Sicherung sämtlicher Publikationen und Digitalisate ist über die Universitätsbibliotheken beider Trägeruniversitäten gewährleistet." (http://www.edition-topoi.org/publishing_with_us/our-partners; on http://www.edition-topoi.org/, Open Access & high quality BoD; Data Repository "Citable", one DOI per object http://www.edition-topoi.org/publishing_with_us/citable [in the medium run also for non-Topoi Members])

> *The "Linked Ancient Data" cloud :* Ancient data are a close ensemble, with precise chronological (and geographical) boundaries. The spider's web works as a methodological paradigm for other worlds.

Editing and collaborative practice:

> *Collaborative editing:* on xml files, but with nice editors and interfaces (more precision, more uncertainty, more transparence of choices, more structure, clarity about data, coherence, control, higher quality)

> *Evaluating and sharing users' contribution with existing sharing systems (Flickr, collaborative platforms)*

> *For every project a big question is: Who does What?* How do we evaluate, assess, stimulate collaboration, and how do we record the intellectual process behind every decision?

The issue of the *traceability of the intellectual journey in order to preserve it in time*

> *How can we validate data coming from digital publications?*

> *How do we assess/evaluate/attribute data?* For example for tenure, etc. And how does it work in different academic environments and countries with different rules?

We need guidelines for evaluating digital scholarship. (see American Historical Association Guidelines: https://goo.gl/zdSdx3).

We need a tool which produces publications that can be traced and attributed.
TEI to critical edition (PDF) http://lombardpress.org/seamlessly-converting-a-tei-critical-edition-to-camera-ready-print-proofs/

One method: how can I contribute, how people can integrate their current work in a larger community?

One training: EpiDoc training and TEI training are widely supported, no more time impact on work to get used to a data entry form but pure coding on the actual epigraphic source.

> Not only one portal but multiple diverse resources and apps based on curated data (the experts' job); encourage developers to do their job and support people to reuse data in their research.

Make data available!

> *How do we computationally harmonize different encoding standards? Up to what stage is EpiDoc "standardized"? And is it even possible to define a standard?*

Chronological and geographical information on inscriptions:

> *How to tag chronological and geographical data within the text in a consistent way?*

> *The material makes its own demands to the editor: chronological data, how they are reported in the inscription? - what do they really mean?*

> *Variety issues:* Geographical information mentioned in epigraphic texts can have an extreme variety in terms of expression - the only thing that they seem to have in common is the fact that the places mentioned are different than the findspot. Various types of chronologies – under specification of the information (month, day, but not the year). Cross-references in inscriptions (both chronological and geographical)

> *We need standardised dating criteria:* choosing periods unanimously accepted by the community; eliminating all references to periodizations based on aspects of art history or on other phenomena that have nothing to do with socio-economic or political changes.

How do we represent chronological data? How do we get closer, not to the modern way of periodizing, but to the proper ancient way of indicating time in ancient sources?

> *How do we represent vagueness?* Something like "around the century ...", "a few years after the battle ...", "close to", "near", etc. - and is it even possible?

> *Standards issues:* How can we encode inscriptions where a fundamental component of geographical information is reported in the form of drawings? Forma Urbis Romae, Water distribution among land-owners, etc.

A strategy: if we start from fixed point, one may wonder whether we can reach to a closer dimension to the actual chronology of the ancient world, without superimposing our scholarly modern assumptions about periods (the example of PeriodO): a suggestion may be a bottom-up approach that has to be adopted in Epigraphy and other fields (Trismegistos is an example of massive manual, easy and bottom-up work within a gazetteer).

How do we link these data together?

Connections between gazetteers and project should be expressed via a standardized protocol. Each publication should make clear how it fits into the eco-system.
Training sessions should be a part of any new project - COST actions? http://www.cost.eu/
Training requires external funding; universities cannot be relied upon to deliver it.
Digital Curation of Cultural Heritage?

Innovative Training Networks (ITN): http://ec.europa.eu/research/mariecurieactions/about-msca/actions/itn/index_en.htm

Presentations and extensive notes:

Pietro Liuzzo:
https://prezi.com/tcuigcqtitne/eagle-and-geo-chrono-data-leipzig-22-april-2016/

An "epigraphy.info" for searching and seeing aggregated content, for editing the source and reviewing with a peer-review workflow (as well as in papyri.info) - we want many editions of the same inscription!

Overview tools: something like Recogito telling what is already there, what is missing - limitate efforts, maximise expertise, offer clarity about coverage, support future projects in their decisions.

We need inventory numbers for the collections! But for that it is very important to get data from museums and curatorial environments.

> *We need to think about how to integrate new systems with "old" museal and archival systems which are nevertheless still used for conservation* - the people who take care of inscriptions need to feel welcome in order to be involved in a way that they can understand. We should start from the big collections so that minor collections feel that they can participate in the "prestigious group".

> *Collaborative editing:* on xml files, but with nice editors and interfaces (more precision, more uncertainty, more transparence of choices, more structure, clarity about data, coherence, control, higher quality).

Common workflow with common tools to do and view things.

One method: how can I contribute, how people can integrate their current work in a larger community.

One training: EpiDoc training and TEI training are widely supported, no more time impact on wok to get used to a data entry form but pure coding on the actual epigraphic source.

> Not only one portal but multiple diverse resources and apps based on curated data (the experts' job); encourage developers to do their job and support people to reuse data in their research.

Make data available!

Silvia Orlandi (Università La Sapienza Roma): *Chronological and geographical information in Latin inscriptions: examples and issues*:

> *Variety issues:* Geographical information mentioned in epigraphic texts can have an extreme variety in terms of expression - the only thing that they seem to have in common is the fact that the places mentioned are different than the findspot. Various types of chronologies - underspecification of the information (month, day, but not the year). Crossreferences in inscriptions (both chronological and geographical).

> *Standard issues:* How can we encode inscriptions where a fundamental component of geographical information is reported in the form of drawings? Forma Urbis Romae, Water distribution among land-owners, A bronze "catasto" from Verona.

The formal external aspect of the chronology requires a different way of encoding information.

> *How do we think to encode these data in the framework of a research tool?*

> The answer depends on the questions that we want to ask: being these questions virtually infinite, encoding should be reduced to a minimum, but the encoded information should be more explicit in order to be re-used.

Anita Rocco (Università di Bari): ***Remarks about time and places in the inscriptions by Christians in Rome***

EDB: the old workflow for annotating geographic data: % for toponyms.

What can we annotate?

- Non Roman origin of the deceased

- Mention of specific places in Rome

- Location of the tomb - micro-regional

Indication of the origin: adjectives, regio, provincia, region, village, name of the location + ex

- Agio-toponyms: introduced by specific adpositional clauses

> A network of *loca sancta* and relevant places in Christian Topography?

Martin Scholz (Friedrich-Alexander Universität Erlangen-Nürnberg): ***EDEN, An Epigraphic Web Database of Ancient Inscriptions***

WissKi ontology's layercake - semantic modelling between objects, places, etc. which are tagged and extracted as rdf triples.

Pathbuilder: semantic backend - paths can be grouped according to typologies.

Graphical editor for annotating texts - places, people, names term, dating - character level (whatever categories that are defined).

Marie-Claire Beaulieu and Tim Buckingham, (Tufts University): ***Classroom epigraphy assignments with Perseids***

> Evaluating and sharing users contribution with Flickr

Silvia Evangelisti, Università di Foggia: ***EDR - Epigraphic Database Roma***

http://www.edr-edr.it/default/index.php

> *We need standardised dating criteria:* choosing periods unanimously accepted by the community; eliminating all references to periodizations based on aspects of art history or on other phenomena that have nothing to do with socio-economic or political changes.

It is necessary to refer to established entries for manual processes - TMGEO

> *Tagging chronological and geographical data within the text?*

Crossreferences to Pleiades and other gazetteers, and establishing a Regulated Vocabulary for Chronological data.

> *How to deal with the enormous amount of time that this operation requires?*

Charlotte Roueché (King's College, London): ***Making the stones speak***

XML and its original benefits: exchange, search, user freedom. However, the benefits were largely internal: not because we didn't want to link, but because there was relatively little to which we could link.

> *The material makes its own demands to the editor: chronological data how they are reported in the inscription - what do they really mean?*

The future will see many specialist gazetteers, each with its own raison d'etre.

> *For every project a big question is: Who does What?*

What questions come next? - Places? People? Events? Dates?

The "Linked Ancient Data" cloud: Ancient data are a close ensemble, with precise chronological (and geographical) boundaries. The spider's web works as a methodological paradigm for other worlds, and Linked Time, instead of Linked Places, might be a good model, because it goes beyond the geographical paradigm of the post-Roman world.

For the Medieval world, of multiple cultures, Linked Time could work the transformation

Michèle Brunet (Université Lyon 2): ***IG Louvre: developments and issues - A case study***

The Theoroi List: the students will work on the inscriptions in the Museum, observe the *topoi* of display in the agora and around, and work on the text and the artifact, as well as the IG XII.

> *How to encode the chronology of the Theorodokoi?* Chronological anchor: a name of a historical person to which a date can be connected.

> *How to visualize the data in order to understand what you are studying and what you are editing?*

TimeMapper might be a first way of visualizing the succession of the data, grouped by three (Workflow in Perseids with students, through a fill-in Google spreadsheet with data; data are imported in Perseids and serialized in a more stable format and made interoperable; afterwards they are visualized in data and links and tags, linking them with other databases in the Open Annotation model).

Second experiment: TimeMapper vs LGPN as crashtest for verifying the result:

> *How can we update the resources and the gazetteers?* They HAVE to be improved because the current scientific work has to correct existing data! Maybe we need Linked Open Data in order to make also old data (old editions in print) available so that they do not get lost.

> The issue of the *traceability of the intellectual journey in order to preserve it in time.*

> *How can we validate data coming from digital publications?*

Monica Berti (Universität Leipzig): ***The Digital Marmor Parium***
Chronological issues: different sources (various lists of archons?)

Shift from an exclusive to an inclusive counting ("around ca. 400 BC" ...)

> How do we represent disagreement between sources and scholars?

Francesco Mambrini and Philipp Frank (Deutsches Archäologisches Institut Berlin): ***Telling stories with the inscription. The EAGLE Storytelling App and beyond***
Having fun with inscriptions

Eagle: some key-concepts

Reach out for the unexplored areas; harmonize, harmonize, harmonize ...; take what is already around and make it available to users; inscriptions are more than a bunch of text lines; reach out for those that are NOT your average epigraphy enthusiasts.

Take inscriptions as objects and allow users to see them, make references bringing together all the best tools available to ancient historians; allow users to build a narrative across inscriptions, embedding resources.
Epidoc Converter API - Philipp Frank

> *How do we computationally harmonize different encoding standards? Up to what stage is EpiDoc "standardized"? And is it possible to define a standard?*

Federico Aurora (University of Oslo): DĀMOS - Database of Mycenean at Oslo

https://www2.hf.uio.no/damos/index/about

1. Stage: an epigraphically annotated corpus

Syllable type - syllabogram-basic, with transliteration, uncertain readings, numberings, conjectured syllables.

Wordtype: logograms, common words, abbreviation, etc. - It is possible to look for words and kinds of words.

2. Stage: Linguistically annotated corpus

- Create an annotation interface and routines: MS access, MySQL Workbench

- Linguistic Annotation (in progress)

- Creating an interface for internet publication

> The problem with linguistic annotation: occurrences connected to various readings

3. Stage: chronology and geography

> How do we record information from the tablets? Relative positions, mentions of regions, internal geographical references > *How to formalize information?*

Francesco Mambrini and Philipp Frank, DAI Berlin: *Data and standards in the iDAI.world*

The Hard data layer: a minor "solar system" of types of data.

The standards: gazetteers, authority files, vocabularies ... iDAI vocab, iDAI.gazetteer, iDAI.chronontology.

What do we have to offer?

- Linkdness within and without DAI's world

- Public APIs to interrogate and get data

- Classics within the world of archaeology

iDAI.vocab: thesaurus of archaeological terminology - it collects technical German terms for archaeology linked to translations into almost 13 languages.

> Public APIs !

iDAI.journal

journals.dainst.org with OAI PMH interface (Open Archive Initiative Protocol for Metadata) to query the collection and harvest metadata.

Text extraction > NLP and open annotation! Key-word extraction! NER, then all back to the open content platform.

iDAI.chronontology

A gazetteer for time definition

connect different definitions for the same period

give chronological coordinates

give geographic coordinates

Frank Grieshaber, Heidelberger Akademie der Wissenschaften and Universität Heidelberg: GODOT: *Graph of Dated Objects and Texts*

Chronology as the backbone of writing history

> Building a chronological gazetteer for antiquity > not much has been done with the aim of harmonizing calendar dates.

> Need for chronological standards in the DH and ancient calendar dates.

A lot of guidance is missing for providing a way of harmonizing data.

Each instance of every calendar entity will have its own URI.

A chronological gazetteer and a research infrastructure.

Stable uris, links between data from various data sources, searching and browsing for dates in the shared data pool, date conversions into Julian calendar and different calendar systems, public API, web annotations as community building feature.

Herbert Verreth, Leuven University: *Trismegistos Places: a geographical index for all Latin inscriptions*

http://www.trismegistos.org/geo/index.php

FileMaker Pro 14.0.1 for an online MySQL / PHP environment.

Capital Clusters: every word starting with a capital listed in the full corpus - if several words started with a capital followed one after the other, they became a capital cluster.

Tagging the capital clusters every word is automatically labelled.

Also other words starting with a capital were found, like mere incipits, names of gods, etc.

Geographical corpus: Egypt, EAGLE, Itinerarium Antonini, Tabula Peutingeriana all tagged

Automatic matching?

Toponyms resembling personal names, difficult strings ... automatic matching doesn't work. Manual intervention.

Phase 1: identification of toponyms in the capital cluster strings

> *Toponyms as spatial entities, in the largest sense*

2. incorporation of the capital cluster into the real TM Georef file

3. The context of the toponym

Toponyms existing out of several consecutive elements are automatically grouped in one card

TM Georef: attested in Latin documentary texts, Latin literary texts

> *How is this annotated and connected to the context of the toponym on the computational side?*

> How to compare various ways partners edit their own material? How to do it automatically?

GeoRef card:

Link to the source, but text also immediately available as the source has it.

Status - explicitly indicated in the text!

Admin_sit - administrative situs explicitly indicated in the text.

Detail: what is relevant for the toponym is indicated in translation as it is in the inscription.

Dates:

They are not always available even in the original source!

> *Please be consistent in new editions!*

Research options:

PlaceNames, toponyms, mentions of toponyms in one text, *ethnika*.

> Updates through swap across databases

Further discussion:

http://ancientgraffiti.wlu.edu/
OpenStreetMap of Herculaneum by AGP:
https://www.openstreetmap.org/#map=19/40.80597/14.34756

Texts are not data. We'd say: texts are not *just* data.

Google Discussion Group "ENcoding COmplex Writing Systems" (ENCOWS)
https://groups.google.com/forum/?hl=de#!forum/encows
Set up after the workshop http://mayawoerterbuch.de/?p=6551&lang=en

PROGRAMME (October 5-6, 2015) [not available online anymore]
Monday, October 5, 2015
13.00 Gabriel Bodard (King's College London)
Ancient Greek and Latin texts, The EpiDoc Community: Training, Infrastructure and the Future
13.30 Thomas Kollatz (Steinheim-Institute for German-Jewish History, Essen)
Jewish funeral inscriptions, Epidat - Hebrew Inscriptions Online
14.00 Max Grüntgens (Deutsche Inschriften Online, Akademie der Wissenschaften Mainz)

German Inscriptions Online, EpiDoc and Web APIs – Ideas, Concepts, Challenges
14.30 Daniel Werning (Humboldt University Berlin, TOPOI)
Egyptian Hieroglyphic Texts: Parameters for a Potential TEI XML Encoding Egyptian Hieroglyphic Writing
15.30 Project Text Database and Dictionary of Classic Mayan (University of Bonn)
Mesoamerican Writing Systems, Maya Hieroglyphic Writing
16.00 Gordon Whittaker (University of Göttingen)
16.30 Annick Payne (University of Basel, EIKONES)
The Palaeography of Anatolian Hieroglyphic Stone Inscriptions
Mesoamerican Writing Systems, Aztec Anatolian Writing Systems, Hieroglyphic Luwian
17.30 Miguel Valério (Universitat de Barcelona)
Linear A and Cypro-Minoan: Writing Strategies and Legibility of Two Mediterranean Writing Systems, Linear A and Undeciphered Syllabaries Cyprominoan
18.00 Hubert Mara (University of Heidelberg)
Cuneiform Tablets & Fingerprints -- Forensic Methods for 3D Data